

DeepSMOTE: Fusing Deep Learning and SMOTE for Imbalanced Data

Year: 2021 | Citations: 390 | Authors: Damien Dablain, B. Krawczyk, N. Chawla

Abstract

Despite over two decades of progress, imbalanced data is still considered a significant challenge for contemporary machine learning models. Modern advances in deep learning have further magnified the importance of the imbalanced data problem, especially when learning from images. Therefore, there is a need for an oversampling method that is specifically tailored to deep learning models, can work on raw images while preserving their properties, and is capable of generating high-quality, artificial images that can enhance minority classes and balance the training set. We propose Deep synthetic minority oversampling technique (SMOTE), a novel oversampling algorithm for deep learning models that leverages the properties of the successful SMOTE algorithm. It is simple, yet effective in its design. It consists of three major components: 1) an encoder/decoder framework; 2) SMOTE-based oversampling; and 3) a dedicated loss function that is enhanced with a penalty term. An important advantage of DeepSMOTE over generative adversarial network (GAN)-based oversampling is that DeepSMOTE does not require a discriminator, and it generates high-quality artificial images that are both information-rich and suitable for visual inspection. DeepSMOTE code is publicly available at <https://github.com/dd1github/DeepSMOTE>.